

Representação lógica do conhecimento para ontologias: um guia para formalização de definições de linguagem natural para lógica descritiva

Guilherme Francis de Noronha¹

Resumo: Ontologias são artefatos usados por profissionais da Ciência da Informação para fins de representação em contextos digitais. Esses artefatos têm se destacado por sua capacidade de representação, de relacionar as entidades e serem processáveis por computadores por meio de lógica descritiva. No entanto, ainda são poucas as ontologias desenvolvidas por pesquisadores da Ciência da Informação que fazem uso da lógica descritiva. Nesse contexto, o presente artigo tem como objetivo familiarizar o profissional da Ciência da Informação com o uso da lógica. Para isso é apresentado os conceitos da lógica descritiva, usada em ontologias, e como elas atuam nos modelos de representação do conhecimento para validação do modelo e geração de conhecimento novo. Como contribuições o presente artigo apresenta um modelo para formalizar a linguagem natural em lógica e apresenta dez orientações de boas práticas para incorporação da lógica em ontologias. Por fim sugere-se a integração dessas orientações numa metodologia de construção de ontologias para consolidar o uso de lógica dentro do contexto da ontologia aplicada.

Palavras-chave: lógica descritiva; ontologia aplicada; Ciência da Informação; formalização do conhecimento.

Logic knowledge representation for ontologies: a guide for formalization of natural language definitions to descriptive logic

Abstract: Ontologies are artifacts used by Information Science professionals, to represent knowledge in digital platforms. Those artifacts are shining due to its representation features like entity relationships and using descriptive logic to be machine-readable. However, there are few ontologies made by Information Science professionals that use descriptive logic. To address this context, the present paper has as objective, familiarize the Information Science professionals with logic. To achieve this, the concepts of descriptive logic is presented to the reader and how it is linked with knowledge representation models to validate models and to reason

¹ Doutorando em Gestão e Organização do Conhecimento pela Universidade Federal de Minas Gerais. E-mail: guilhermenoronha2001@gmail.com. Lattes: <http://lattes.cnpq.br/9884915193147340>



new knowledge. The contributions of this paper is to provide a model to formalize natural language in logic and provide ten guidelines of how to incorporate logic in ontologies. Lastly, this paper suggests as future work to embody those guidelines into an ontology methodology as the best way to consolidate the use of logic within the applied ontology context.

Keywords: descriptive logic; applied ontology; Information Science; knowledge formalization.

1 INTRODUÇÃO

As ontologias desempenham papel importante nos sistemas modernos de informação que precisam de interoperabilidade semântica e representação do conhecimento. É uma área interdisciplinar que se integra, principalmente, nos campos da Filosofia, Ciência da Informação (CI) e Ciência da Computação (CC).

Segundo Almeida (2014), existem correlações para o uso de ontologias nessas três diferentes áreas do conhecimento. A ontologia no campo da filosofia é uma disciplina sobre sistemas de categorias que busca entender a realidade, as coisas que existem e suas características. Na CI e na CC, as ontologias são consideradas artefatos. Na CI, elas são usadas como teorias informais para entender domínios e classificar sujeitos e, também, como um sistema conceitual informal para criar vocabulários controlados que serão usados em sistemas de recuperação de informação. Na CC, os artefatos são produtos de softwares usados para criar vocabulários para representação de sistemas e, também, uma teoria baseada em lógica usada para entender um domínio, modelá-lo e gerar inferências.

No contexto digital atual as noções para ontologia em CI e CC se sobrepõem, e as teorias desenvolvidas num campo são aplicadas na prática de outro. As ontologias passam a ser, necessariamente, artefatos computacionais que são desenvolvidos baseando-se em teorias de representação da informação da CI.

A adoção do aparato tecnológico no desenvolvimento de ontologias ganhou força ainda na década de 90, com a popularização dos computadores pessoais, a ascensão das páginas web e sua incorporação no escopo da CC (VICKERY, 1997). As ontologias, então, ganharam contribuições da CC para que os modelos de representação da informação passassem a ser processáveis por computadores por



meio de famílias de lógicas de conhecimento tais como a lógica descritiva (BAADER et al., 1990).

A principal delas é a Web Ontology Language (OWL), linguagem com semântica bem definida para representação do conhecimento via ontologias (PATEL-SCHNEIDER et al., 2004). A OWL é fundamentada na evolução da lógica descritiva, área do conhecimento que intercala capacidade de processamento computacional com representatividade e expressividade semântica. Por meio dessa linguagem é possível construir ontologias e disponibilizá-las de forma que possam ser reusadas e incorporadas em sistemas de organização do conhecimento. A lógica descritiva permite que inferências sejam realizadas e novos conhecimentos sejam gerados, além de permitir consultas sobre as entidades representadas.

A OWL, assim como outras tecnologias, é uma linguagem acessível para todos os públicos e dispensa um especialista em tecnologia. Ela é padronizada pela W3C², comunidade internacional, que desenvolve padrões para tecnologias web. A linguagem é, portanto, plenamente documentada e adotada por profissionais da informação ajudando, assim, na disseminação da lógica em ontologias.

Dessa maneira, várias metodologias e ferramentas de interface gráfica surgiram numa tentativa de popularizar o desenvolvimento de ontologias. Um estudo realizado por Mendonça (2015) levantou metodologias de construção de ontologias entre elas a Methontology (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997), 101 (NOY; MCGUINNESS et al., 2001), NeOn (SUÁREZ-FIGUEROA et al., 2012), dentre outras.

Fernández, Gómez-Pérez e Corcho (2004) já destacavam a ausência de uma metodologia unificada de ontologias, cenário este que vem ganhando força (SANTOS, 2014). Um resultado, no contexto da CI, foi a criação da OntoForInfoScience (MENDONÇA; ALMEIDA, 2016), uma metodologia voltada para profissionais da Ciência da Informação. Segundo os autores, a proposta dessa metodologia é possibilitar que os profissionais da Ciência da Informação superem

2 World Wide Web Consortium. Disponível em: <https://www.w3.org/>.



os jargões técnicos, questões lógicas e filosóficas envolvidas da criação de ontologias.

No entanto, uma análise das metodologias apresentadas, revelou uma ausência de definições técnicas e diretrizes para formalização das definições representadas pelas ontologias. Seppälä e Ruttenberg (2013) corroboram com essa dificuldade ao citar a ausência de uma metodologia para formalização em ontologias.

A importância do uso da lógica atrelada ao campo da ontologia aplicada³ é ressaltada por autores seminais de ontologias na CC e na Filosofia, como Guarino (1995), Gruber (1995) e Smith (2003). Os autores identificam como benefícios imediatos do uso da lógica em ontologias: (1) formalização da linguagem natural; (2) processamento por computadores; (3) precisão da recuperação da informação; (4) inferências de conhecimentos novos, dentre outros. Seppälä, Schreiber e Ruttenberg (2014) complementam informando que as definições formais ajudam a eliminar a ambiguidade das definições dos conceitos em linguagem natural.

Destaca-se que as contribuições que as ontologias têm a oferecer no campo da CI e CC passam pela representação do conhecimento no ambiente digital, visto o fenômeno, que vive-se há mais de 30 anos, da explosão informacional disponível.

Nesse contexto, o presente artigo busca popularizar o uso da lógica para a formalização de definições em linguagem natural usada em ontologias. O **objetivo geral** é apresentar e compilar diretrizes para a formalização em lógica de definições em linguagem natural em ontologias, tornando o processo acessível ao profissional da Ciência da Informação. Para tal, busca-se fundamentar os conceitos básicos da lógica para o entendimento e desenvolvimento da formalização durante a construção de ontologias; correlacionar a lógica descritiva, que é processável por computadores, com a ontologia aplicada e; apresentar exemplos de formalização como um guia de boas práticas.

3 Ontologia aplicada é um ramo da filosofia aplicada que usa os conceitos filosóficos e métodos de ontologias para contribuir com a representação do conhecimento científico (MUNN; SMITH, 2013).



Justifica-se o esforço pela necessidade de explorar as funcionalidades que as ontologias oferecem quando definidas formalmente por meio de lógica. Tais benefícios vêm sendo discutidos entre ontologistas e profissionais da Ciência da Informação. Projetos renomados, como a Gene Ontology (GO), esforços vêm sendo feitos para a formalização da ontologia em lógica após as definições formais em linguagem natural (Gene Ontology Consortium, 2015). Acredita-se que tal esforço poderia ter sido minimizado se a formalização já estivesse presente durante a criação da ontologia.

O presente trabalho está organizado da seguinte maneira: a seção 2 contextualiza os fundamentos da lógica e esclarece sua importância para a ontologia aplicada; a seção 3 explica o que são e como são feitas as inferências em lógica descritiva, bem como o uso com as ontologias; a seção 4 introduz a simbolização para formalização da linguagem natural para a lógica; a seção 5 compila diretrizes e orientações, num total de dez, para a formalização da linguagem natural para a lógica; a seção 6 exemplifica a formalização na prática; a seção 7 traz as discussões e trabalhos futuros e; por fim a seção 8 traz os agradecimentos.

2 CONCEITOS BÁSICOS DA LÓGICA

Uma definição adequada para a **lógica** é dada por Genesereth e Kao (2013) que a define como o estudo da informação codificada na forma de definições lógicas. De forma básica, qualquer lógica consiste de uma linguagem mais um sistema dedutivo. Em geral, a **linguagem** é um subconjunto da linguagem natural, como o português ou o inglês. O **sistema dedutivo** objetiva capturar as inferências corretas para certa linguagem. As **inferências**, por sua vez, são deduções de conhecimento novo a partir de conhecimento existente.

A ontologia aplicada tem como base as definições da lógica aristotélica e seu sistema de categorias. Para Aristóteles (2016), as definições formais podem ser dadas usando-se os conceitos de sujeito, cópula e predicado, que definem a estrutura das formalizações. O **sujeito** é o assunto sobre o qual é tratada a definição e pode ser dividido em universal, particular e singular como demonstrado no Quadro 1. O **predicado** é o conjunto de características que definem o sujeito. Por fim, a **cópula** é a conexão entre sujeito e predicado



(BALLARD; COPI, 1978). A estrutura simples de uma definição é dada pela seguinte estrutura: “*Sujeito é Predicado*”.

Quadro 1 – Julgamento proposicional do sistema aristotélico

Universal	Particular	Singular
Todos os sujeitos	Algum sujeito	Esse Sujeito
Todos os cães possuem pelos	Algum cão possui pelos	Rex possui pelos

Fonte: Almeida (2014) adaptado pelo autor.

O sujeito, predicado e cópula são usados para a construção de definições formais em lógica. Essas definições, por sua vez, podem formar **argumentos** que é a justificativa que acompanha as premissas por meio de uma conclusão. As **premissas** são definições usadas num argumento para defender uma conclusão. Por fim, a **conclusão** é o resultado que se espera quando as premissas defendidas são verdadeiras (DIMNET, 2020). Esses conceitos são importantes para o entendimento de inferências na lógica.

Uma **definição lógica** é o conjunto de, pelo menos, um sujeito conectado por, pelo menos, um predicado e podem ser simples ou complexas. Essas conexões entre sujeito e predicado são feitas por meio de operadores lógicos que fazem o papel da cópula. As definições simples são compostas de apenas um sujeito e um predicado; e as compostas contêm dois ou mais sujeitos e predicados conectados por dois ou mais operadores lógicos. O resultado desse conjunto possui um valor, podendo ser verdadeiro, falso ou desconhecido.

As coisas existentes podem ser classificadas de acordo com a sua essência, que são conjuntos de características que descrevem unicamente um conjunto do mesmo tipo. A distinção de Aristóteles é chamada de gênero-espécie que pode ser descrito como sujeito no sentido geral e o predicado como aquilo que o diferencia dos demais sujeitos. Sujeito e predicado são englobados pelo **universo do discurso**, que é o conjunto de entidades relevantes em que o modelo de representação do conhecimento é baseado (USÓ-DOMÉNECH; NESCOLARDE-SELVA; GASH, 2018).



As ideias de Aristóteles contribuem para as ontologias de duas maneiras: (1) dão suportes para o realismo filosófico necessário na construção de ontologias cujo o domínio se baseia em conceitos universais (SMITH; CEUSTERS, 2010) e; (2) fundamentam o conceito de lógica clássica e traz as primeiras ideias sobre inferências por meio de lógica.

O **realismo filosófico** estabelece uma completa independência da realidade ontológica em relação as concepções formadas em torno dos conceitos. Ou seja, um conceito existe independente do sujeito que o observa. O realismo não é a única abordagem filosófica em ontologias. No entanto, ela se mostra como a mais adequada para representação do conhecimento científico, uma vez que o conhecimento representado são independentes de pontos de vista e outros fatores externos (ARP; SMITH; SPEAR, 2015).

A lógica Aristotélica, no entanto, era insuficiente para cobrir todo o raciocínio lógico que a dedução exigia. Foi somente no século XX que foi formulada uma teoria explícita, sobre inferências, que era adequada para explicar o raciocínio dedutivo na matemática e ciências empíricas (SUPPES, 2012).

A lógica moderna se estabelecia e mais tarde seria a responsável pela criação da lógica de programação, como um braço da filosofia que se desenvolveu em campo computacional. Nota-se então que a lógica faz parte das ontologias tanto do ponto de vista filosófico, quanto do ponto de vista matemático.

Existem vários tipos de lógica, em geral, organizadas em duas categorias principais: clássica e não-clássica (BALLARD; COPI, 1978). Cada um desses tipos envolvem variações, gerando dezenas de famílias de lógica. Um **tipo de lógica** pode ser diferenciado pela sua capacidade de representação. Ou seja, sua expressividade em relação às definições que descrevem os fatos.

A **expressividade** é a capacidade de uma linguagem definir um sujeito formalmente. Podemos considerar a linguagem natural como a forma mais expressável de definir um sujeito. A lógica, conforme utilizada na filosofia, manipula a linguagem natural. Ambas não são processáveis por computadores, pois elas podem ser ambíguas enquanto os algoritmos, base dos sistemas de



informação, não permite esse tipo de indecidibilidade (SIPSER, 1996). No mundo digital, quanto maior a expressividade da lógica, menor seu poder computacional.

Primeiramente é preciso explicar algumas definições comuns em lógica, como lógica de primeira ordem, predicados, quantificadores e operadores lógicos. As definições foram retiradas de Gensler (2010), mas podem ser obtidas em qualquer livro de lógica para entender linguagens formais.

Para entender predicados é necessário analisar uma definição que contenha um sujeito e um predicado: “o cão late”. O sujeito, que é particular, é “cão” e o predicado é “late”. Note que o predicado é sempre algo que diz respeito ao sujeito. Para formalizar essa definição atribuímos ao sujeito uma letra minúscula para simbolizar uma constante, como a letra “c” para cão. Para o predicado atribuímos uma letra maiúscula para simbolizar uma função, como a letra L para “x late”, onde “x” é um sujeito universal. Por fim, podemos formalizar a definição “o cão late” como “Ld”.

Os **operadores lógicos**, assim como os operadores aritméticos da matemática, são utilizados para fazer operações entre um ou mais sujeitos. Os resultados entre essas operações é um valor binário que pode representar verdadeiro ou falso. Geralmente esses valores são definidos por tabelas verdade⁴ que determinam a validade das formalizações (GENESERETH; KAO, 2013).

Aqui cabe um parêntese para ressaltar a diferença entre verdade e validade na lógica. A **validade** diz respeito aos argumentos. Um argumento é válido se as conclusões são justificadas pelas premissas. Considere a definição “Se o sujeito é um cão, logo ele é mamífero” formalizado por “ $Cx \rightarrow Mx$ ”. Se “x” é uma instância de “cão” então “Cx” é verdadeiro e “Mx”, por consequência, também é verdadeiro. Nesse exemplo, o resultado da operação “ $Cx \rightarrow Mx$ ” deverá ser necessariamente verdadeiro. Já a **verdade** tem a ver com as premissas. “Ser cão” pode ser verdadeira ou falsa dependendo de qual sujeito ela está relacionada. No mundo real, por exemplo, “o homem é cão” é uma premissa falsa.

⁴ Tabela verdade é um conjunto de linhas e colunas que relacionam um ou mais sujeitos e seus valores de verdade. É utilizada pelos lógicos para determinar se uma expressão é válida.



Quanto aos quantificadores existem dois: universal e existencial. O **quantificador universal** é dado pelo símbolo \forall e pode ser lido como “Para todo x ” e significa todos. O **quantificador existencial** é dado pelo símbolo \exists e pode ser lido como “Existe um x ” e significa algum(ns). Numa formalização o quantificador universal pode ser usado da seguinte maneira: “ $\forall x (Cx \rightarrow Mx)$ ”. Essa formalização pode ser lida da seguinte maneira: “Para todo x , se x é um cão, então x é um mamífero”. De forma similar o quantificador existencial é usado, como por exemplo: “ $\exists x (Mx \wedge Lx)$ ”. Essa formalização pode ser lida como “Existe um x tal que, x é um mamífero e x late”.

Por fim, podemos definir uma linguagem **lógica de primeira ordem** como sendo um sistema de lógica simbólica que usa predicados, quantificadores e operadores lógicos. Além dessas considerações básicas, cabe destacar que a linguagem mais utilizada para a construção de ontologias é a OWL. A **OWL** é uma linguagem de marcação semântica usada publicar e compartilhar ontologias na web (MCGUINNESS, HARMELEN, 2004). Algumas sublinguagens da OWL, como a OWL-DL podem ser derivadas em lógica descritiva que é um subtipo da lógica de primeira ordem apresentada acima.

As **lógicas descritivas** são usadas para formalizar conteúdos de domínios de forma estruturada (BAADER; HORROCKS; SATTLE, 2009). Elas representam um meio termo entre a expressividade, necessária para representar domínios do conhecimento, e processamento por computadores, requisito para uso em sistemas de informação. A Tabela 2 mostra os operadores suportados pela lógica descritiva e pela OWL-DL (BECHHOFFER et al., 2004).

Tabela 1 – Operadores suportados pela lógica descritiva e pela OWL-DL.

NOME DO OPERADOR	DESCRIÇÃO	REPRESENTAÇÃO EM LINGUAGEM NATURAL	REPRESENTAÇÃO SIMBÓLICA
Negação	Nega uma definição	“Platão não é imortal”	$\neg Ip$
Conjunção	Interseção entre duas definições	“Platão é homem e Platão é mortal”	$Hp \wedge Mp$
Disjunção	União entre duas definições	“Platão é mortal ou Platão não é imortal”	$Mp \vee \neg Ip$
Quantificação Existencial	Existe pelo menos uma definição que satisfaz a condição de verdade do conjunto	“ Existe pelo menos um mortal chamado Platão ”	$\exists x: Mx \equiv Mp$



Quantificação Universal	Todas as definições de um conjunto são verdadeiras	<i>“Todos os homens são mortais”</i>	$\forall x: Hx \rightarrow Mx$
Equivalência	Uma definição é equivalente a outra	<i>“Mortal é equivalente a não-imortal”</i>	$M \equiv \neg I$
Indivíduo	Indica que o universo de discurso possui pelo menos um sujeito que satisfaz a condição de verdade	<i>“Toda instância é mortal ou não-mortal”</i>	$Mx \sqcup \neg Ix \equiv T$
Conceito Vazio	Indica que o universo de discurso não possui sujeitos que satisfazem a condição de verdade	<i>“Nenhuma instância é mortal e imortal”</i>	$Mx \cap Ix \equiv \perp$
Inclusão	Um sujeito está incluído dentro de um subconjunto de um predicado	<i>“Platão está incluído dentro do predicado homem”</i>	$p \sqsubseteq H$
Definição	Um sujeito está definido de acordo com um conjunto de predicados	<i>“Platão é definido como homem e mortal”</i>	$p \doteq H \cap M$
Conceitualização	Um sujeito é atribuído a um predicado	<i>“Platão é homem”</i>	$p : H$
Relação	Um sujeito se relaciona a outro sujeito por meio de predicados	<i>“Platão é mestre de Aristóteles”</i>	$(p,a) : M$

Fonte: Genesereth e Kao (2013) adaptado pelo autor.

A representação simbólica da Tabela 1 segue as convenções sugeridas por Suppes (2012). O autor usa letras minúsculas para representar sujeitos e letras maiúsculas para representar predicados. Os valores de “x” são variáveis que dependem do universo de representação.

Nota-se que além dos operadores citados na Tabela 1, a OWL-DL contém outros recursos que vão além da lógica descritiva. Esses recursos, como atributos e restrições, são conceitos adotados em linguagens de programação que ajudam a modelar uma ontologia. As restrições limitam os atributos que um sujeito pode ter. Esses atributos, por sua vez, são tipos de dados que possuem propriedades e que podem ter restrições de cardinalidade. Por exemplo, o sujeito “cão” possui os atributos “raça” e “idade” limitados a um atributo por sujeito do tipo “cão”. O atributo “idade” possui como propriedade ser um número inteiro e sua restrição é que seja um número positivo.

Os operadores listados na Tabela 1 permitem a derivação de diferentes linguagens lógicas que são processáveis por computadores. Para definir a sintaxe que uma linguagem pode suportar, criou-se uma convenção que descreve quais são os operadores permitidos (HARMELEN; LIFSCHITZ; PORTER, 2008). Essas convenções permitem que motores de inferência sejam construídos para validar



um modelo de dados e deduzir conhecimento novo a partir de conhecimento existente. Se um cão é definido com uma idade negativa, por exemplo, um motor de inferência apontará uma inconsistência na ontologia. É exatamente esse o grande diferencial das ontologias para outros modelos de representação do conhecimento.

3 INFERÊNCIAS EM LÓGICA DESCRITIVA: REPRESENTAÇÃO NA ONTOLOGIA APLICADA

Na lógica descritiva são divididas em três partes diferentes: inferências para Tboxes; inferências para Aboxes, e; inferências dos operadores lógicos (GIACOMO; LENZERINI, 1996).

Um **Tbox** é a representação intensional de um conhecimento em forma de terminologia (GRUBER, 2009). Um conhecimento é intensional quando ele fornece o mínimo necessário para que um sujeito saiba diferenciar um conceito de outro (AGANETTE; ALMEIDA, 2015). É construído por meio de declarações que descrevem as propriedades gerais dos sujeitos. As terminologias que formam os Tboxes são formadas pelas junções de sujeitos que já foram previamente definidos. Por exemplo:

(1) Exame hematológico e linfático \equiv Exame hematológico \sqcap Exame linfático (FARINELLI et al., 2016).

Por meio de um Tbox é possível realizar inferências em sujeitos intensionais. Por exemplo, ao definir um “exame” como “*hematológico e linfático*”, logo conclui que ele é a conjunção entre os exames “*hematológico*” e “*linfático*”.

Um **Abox** é formado por atribuições de instâncias aos sujeitos definidos pelas Tboxes (GRUBER; 2009). Por exemplo:

(2) Exame hematológico(velocidade de hemossedimentação) \sqcap exame linfático.

Essa formalização afirma que o exame de “*Velocidade de hemossedimentação*” é um “*exame hematológico*”, pois faz coleta de sangue e também é um “*exame linfático*”, pois é usado para detectar inflamações no linfoma. As Aboxes também fazem atribuições de relações. A formalização 3 ilustra um exemplo de relações em Aboxes:



(3) Dota de ervo(nervo, órgão) (Gene Ontology Consortium, 2015).

A relação “*dota de nervo*” é atribuída de um “*nervo*” para um “*órgão*”. Em linguagem natural pode-se dizer que “*nervo*” dota de nervo “*órgão*” e que “*órgão*” é dotado de nervo por “*nervo*”. Esta última relação possui propriedade inversa da primeira e que pode ser representada em OWL. Inferências em Aboxes baseiam-se principalmente na verificação de instâncias que podem pertencer a um sujeito especificado (BAADER, 2003).

As inferências em operadores lógicos são possíveis graças ao método dedutivo que utiliza-se de argumentos com premissas para chegar a uma conclusão que pode ser verdadeira ou falsa. Por exemplo, as definições “*Todo homem é mortal*”, “*Aristóteles é homem*” e “*Logo Aristóteles é mortal*” são um conjunto de duas premissas e uma conclusão.

Um argumento é válido se sua premissa e conclusão são verdadeiras. A partir disso pode-se deduzir um conjunto de argumentos válidos que são usados para simplificação lógica e dedução automática. Por exemplo, a partir de duas premissas “*Ser humano possui sexo biológico masculino ou feminino*” e “*Maria não possui sexo biológico masculino*” pode-se deduzir que “*Maria possui sexo biológico feminino*”.

A Tabela 2 mostra as principais inferências que são realizadas por meio de operadores lógicos. Elas foram criadas a partir de premissas condicionadas umas as outras, ou seja, se as premissas são válidas, então a conclusão também é válida.

Tabela 2 – As inferências dedutivas da lógica descritiva.

REGRAS DE INFERÊNCIA	OPERADORES	DEMONSTRAÇÃO EM LINGUAGEM NATURAL
Modus ponens	$H \rightarrow M$	Todo homem é mortal. Se sou homem, logo sou mortal.
Modus tollens	$H \rightarrow M$	Todo homem é mortal. Se não sou mortal, logo não sou homem.
Silogismo disjuntivo	$M \sqcup F$	O ser humano possui sexo biológico masculino ou feminino. Se não possuo o sexo biológico feminino, logo meu sexo biológico é masculino.
Adição	$M; F$	O ser humano é do sexo masculino. Se o ser humano é do sexo feminino, logo o ser humano é do sexo masculino ou do sexo feminino.
Simplificação	$A \sqcap R$	O ser humano é animal e racional. Se sou ser humano, logo sou animal e racional.
Conjunção	$A; R$	O ser humano é animal e racional. Se sou ser humano e animal, logo sou racional.

Silogismo hipotético	$H \rightarrow A; A \rightarrow M$	Todo homem é animal. Se todo animal é mortal, logo todo homem é mortal.
Eliminação universal	$\forall x: H(x) \rightarrow M(x)$	Todos os homens são mortais. Se Platão é homem, logo Platão é mortal.
Eliminação existencial	$\exists x: M(x)$	Existem coisas que são mortais. Platão é mortal.

Fonte: Genesereth e Kao (2013) adaptado pelo autor.

As inferências por operadores lógicos se apoiam na validade das premissas. Para que o modus ponens, por exemplo, seja verdadeiro, é necessário que a premissa ao qual ele se baseia seja verdade. Por exemplo, a premissa do modus ponens diz que “Se eu tenho *A*, logo tenho *B*”. Partindo da premissa “*todos os homens são mortais*”, temos que “*A*” é “*homem*”, “*B*” é “*mortal*” então “Se sou homem, logo sou mortal”.

Por exemplo, somente é possível deduzir que ser homem é também ser mortal se a premissa “*todos os homens são mortais*” for verdadeira. No entanto, se a conclusão for falsa, a premissa também deverá ser falsa: “Se não sou mortal, logo não sou homem”. Nesse sentido, ressalta-se a importância para que a modelagem de uma ontologia seja feita usando-se premissas se aplicam para todos os casos que ela representar. Do contrário, as inferências realizadas estarão incorretas.

As inferências por operadores lógicos são usadas de várias maneiras em ontologia aplicada. Na *Basic Formal Ontology* (BFO) (ARP; SMITH; SPEAR, 2015), por exemplo, o sujeito “*Processo*” é uma subclasse de “*Ocorrente*” que é uma subclasse de “*Entidade*”. Um motor de inferência usa o silogismo hipotético para deduzir que “*Processo*” é uma subclasse de “*Entidade*”. O silogismo disjuntivo, por sua vez, é usado para separar sujeitos distintos como “*Ocorrente*” e “*Continuante*”. As eliminações universal e existencial servem para separar os sujeitos universal e particular, respectivamente, em sujeito singular.

Os sujeitos de uma formalização ligada por operadores lógicos estão sujeitos a quatro tipos diferentes de inferência: satisfatibilidade, subsunção, equivalência e disjunção mútua (BAADER et al., 2010).

O conceito de **subsunção**, premissa menor ou do inglês *subsumption*, representado por $C \sqsubseteq D$, é a validação de se o sujeito subsumidor (*D*) é considerado mais geral que o sujeito subsumido (*C*). Em outras palavras, é a verificação se a primeira definição pertence ao conjunto denotado pela segunda.



A **satisfatibilidade** é determinar se uma definição possui o conceito vazio (\perp) ou não. A equivalência verifica se uma definição possui as mesmas propriedades de outra, ou seja, são iguais se tem suas expressões expandidas.

Por último, a **disjunção mútua**, onde duas ou mais premissas só têm conclusão verdadeira quando apenas uma delas for verdadeira. Esta verifica se a interseção de duas definições num dado conjunto é vazia. Em regras gerais toda inferência realizada na lógica descritiva pode ser reduzida a um caso de subsunção. Baader et al. (2010) mostram que dado os sujeitos C e D se tem:

- C é satisfazível se, e somente se, C é subsumido por \perp ;
- C e D são equivalentes se, e somente se, C é subsumido por D e D é subsumido por C;
- C e D são mutuamente disjuntos se, e somente se, $C \sqcap D$ é subsumido por \perp .

As reduções para subsunção permitem que motores de inferência sejam construídos baseando-se nesses conceitos. Esses métodos também podem fornecer, além da satisfatibilidade das expressões, a conjunção e a negação.

A redução dos casos para subsunção levaram a criação de novos métodos para verificação de satisfatibilidade de uma expressão, conhecido como cálculo de *tableau*, que é um método formal de dedução lógica que está presente em diferentes tipos de linguagens lógicas. O *tableau* é utilizado para verificação de satisfatibilidade das definições lógicas. O método foi incorporado em motores de inferência para OWL como o Pellet (SIRIN et al., 2007), HermiT (SHEARER; MOTIK; HORROCKS, 2008) e FaCT++ (TSARKOV; HORROCKS, 2006) e segue constantemente sendo aperfeiçoado (MEHRI; HAARSLEV; CHINAEI, 2018).

4 REPRESENTANDO A LINGUAGEM NATURAL EM LÓGICA

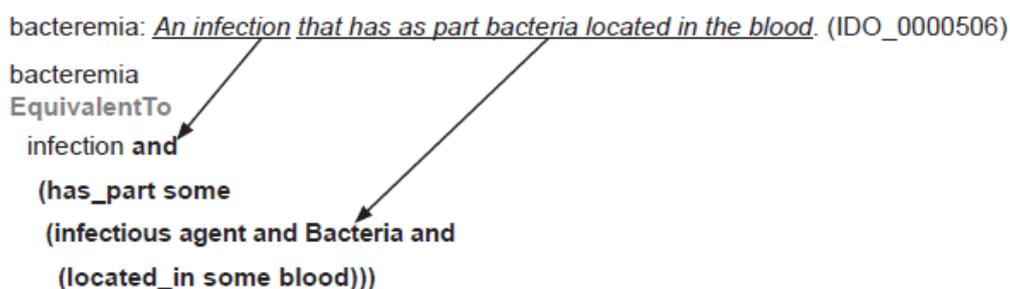
A ontologia aplicada trata de representar o conhecimento humano, que é codificado em linguagem natural. Há, no entanto, uma diferença entre a expressividade da linguagem usada no dia a dia e a simbolização lógica. A lógica simbólica não é capaz de representar substantivos, pronomes, verbos ou adjetivos.



Em geral, as variações gramaticais mais comuns não podem ser distinguidas pela lógica (SUPPES, 2012).

A linguagem natural, no entanto, permite a construção de sentenças da mesma forma que a representação lógica requer. Essa representação em linguagem natural é, inclusive, recomendada por Seppälä, Ruttenberg e Smith (2017). Os autores sugerem um guia para formalizações em ontologias. Esse guia tem como propósito padronizar as formalizações e seguem uma série de princípios tais como, usar apenas um gênero, ter definição única, evitar plurais, evitar repetições, entre outros. As orientações ajudam a escrever boas definições em linguagem natural que são mais facilmente simbolizadas logicamente como mostradas na Figura 1.

Figura 1 – Definição em linguagem natural de *bacteremia* e formalização em OWL.



Fonte: Seppälä, Ruttenberg e Smith (2017)

Smith et al. (2014) apresentaram a teoria da ontologia aplicada para a produção de formalizações do conhecimento. No entanto, tais formalizações esbarram no sentido ambíguo que as palavras produzem na linguagem natural. Durante a aquisição do conhecimento há dificuldades para a tomada de decisões, que se tornam complexas quando não se tem o conhecimento ou o auxílio do especialista da área. Seppälä, Schreiber e Ruttenberg (2014) sugerem que formalizações em linguagem natural em ontologias carecem de formalizações lógicas correspondentes. No entanto, no mesmo trabalho, os autores mostraram que ainda existe a necessidade de metodologias que encubram essa área do conhecimento.

Para representar a linguagem natural em lógica, é preciso, primeiramente padronizar as formalizações. A linguagem natural permite que as formalizações sejam feitas usando-se os mesmos sujeitos, predicados e quantificadores adotados pela lógica.



Nesse sentido, o sujeito torna-se um componente básico da formalização. Pode ser uma variável, um nome próprio ou uma descrição formal. Os sujeitos, sem predicados não possuem valores de verdades associados a eles.

Os predicados, por sua vez, é toda parte estrutural da formalização em linguagem natural que descreve o sujeito. Pode ser composto de outros sujeitos, preposições, artigos, verbos e outras combinações de classes gramaticais disponíveis na linguagem natural. O valor de verdade dos predicados está ligado aos sujeitos que estão relacionados com a sua formalização (SUPPES, 2012). A definição de “*bacteremia*” da Figura 1 é um exemplo de formalização em linguagem natural que utiliza-se de sujeitos e predicados. O sujeito é “*bacteremia*” e o predicado é o resto da sentença que o descreve.

Os quantificadores são construções que especificam a generalização de sujeitos dentro do universo do discurso. Eles se aplicam a, ou satisfazem a, um predicado tornando-o aberto, ou seja, com um valor de verdade para sujeitos universais e particulares. Na lógica, não é preciso substituir os termos para atestar se uma definição é verdadeira ou falsa. Os quantificadores servem para atestar que um predicado é verdadeiro independentemente se os termos forem conhecidos ou não. Na Figura 1, a noção de quantificadores está implícita na sentença e não é demonstrada em linguagem natural. Subentende-se que toda “*bacteremia*” possui os predicados mencionados na formalização. Da mesma maneira, subentende-se que os predicados possuirão algum valor de verdade que corresponde com a formalização proposta. As noções de “todo” e “algum” em linguagem natural corresponde aos símbolos lógicos de (\forall) e (\exists) respectivamente.

Além de identificar os sujeitos, predicados e quantificadores, é preciso também traduzir as noções da linguagem natural que representam os operadores lógicos, restrições e atributos. Para isso, o presente artigo propõe ao leitor a Tabela 3, que é a noção dessas representações para a linguagem natural.

Tabela 2 – As inferências dedutivas da lógica descritiva.

OPERADORES	REPRESENTAÇÃO EM LINGUAGEM NATURAL	COMO IDENTIFICAR O OPERADOR NA LINGUAGEM NATURAL
\equiv	Same, equivalent, equal, etc.	A linguagem indica que uma formalização é equivalente a outra



\subseteq	Subset, part of, including, etc.	A linguagem indica que uma formalização está contida em outra
:	Is a	A linguagem indica que um sujeito está sendo definido por um predicado
\sqcap	And	A linguagem indica que dois ou mais predicados são conjuntamente necessários para que a formalização seja válida
\sqcup	Or	A linguagem indica que dois ou mais predicados são excludentes, ou seja, só pode haver um predicado válido para que a formalização seja válida
\neg	Not	A linguagem nega um sujeito ou predicado
P^{-}	Inverse of, contrary, antonym, etc.	A linguagem indica que uma formalização é oposta a outra
\forall	All, For all, every, etc.	A linguagem indica que o sujeito definido é universal
\exists	Exists, there is, etc.	A linguagem indica que o sujeito definido é particular
$\exists!$	Only one, exactly one, etc.	A linguagem indica que o sujeito definido é singular e só existe um.
=	Equal, even, like, etc.	A linguagem indica uma restrição que é igual a um número real x
>	Greater, more than, bigger, etc.	A linguagem indica uma restrição que é maior a um número real x
\geq	Greater or equal, etc.	A linguagem indica uma restrição que é maior ou igual a um número real x
<	Less, smaller, etc.	A linguagem indica uma restrição que é menor a um número real x
\leq	Less or equal, etc.	A linguagem indica uma restrição que é menor ou igual a um número real x

Fonte: Elaborado pelo autor.

A Tabela 3 serve como consulta para as possíveis representações de operadores lógicos, restrições e atributos. Escolheu-se o inglês para exemplificar as formas de representação da informação pelo fato de ser uma linguagem universal, mas as regras apresentadas se aplicam para qualquer idioma. Além disso, ela não exausta todas as possibilidades da linguagem, apenas apresenta os conceitos para que o profissional da Ciência da Informação saiba identificá-los.

Com base nesses conceitos é possível representar a linguagem natural em lógica. A definição (4), por exemplo, é o resultado em lógica da representação da formalização mostrada na Figura 1. A relação de equivalência (\equiv) indica que o sujeito “*bacteremia*” possui todas as características do predicado que o define. O predicado, por sua vez, é composto de cinco sujeitos diferentes de forma conjunta (\sqcap), pois todos eles são necessários para que a formalização de “*bacteremia*” seja válida. O quantificador existencial (\exists) está conectado aos sujeitos de forma que o valor de verdade seja associado somente por meio da eliminação existencial. Dessa



forma, a formalização em lógica permanece verdadeira, ainda que nenhum sujeito singular esteja formalizado ao modelo de representação do conhecimento.

(4) *“bacteremia”* \equiv *“infection”* \sqcap (*“has part”* \exists (*“infectious agent”* \sqcap *“Bacteria”* \sqcap *“located in”* \exists (*“blood”*)))

5 ORIENTAÇÕES PARA SIMBOLIZAÇÃO EM OWL

As seções 2 e 3 trouxeram a fundamentação da lógica para sistemas de representação de informação como ontologias. A seção 4 trouxe o método para representar a linguagem natural em lógica. Com base nas seções anteriores, é possível formalizar em lógica as formalizações em linguagem natural encontradas em ontologias.

O presente capítulo, portanto, trata de condensar toda a informação apresentada até o momento para elaborar um guia de boas práticas para a formalização lógica, em especial usando a OWL, de linguagem natural.

A literatura apresentada de representação de linguagem natural é adaptada para que os operadores da lógica sejam usados e compreendidos pelo profissional da Ciência da Informação. O objetivo dessa seção é apresentar dez orientações de formalização lógica para definições em linguagem natural em ontologias. Buscou-se traduzir os operadores, quantificadores, atributos e restrições da lógica descritiva e associá-los aos sujeitos da linguagem natural. Essa associação ajudará o profissional da Ciência da Informação a identificar os sujeitos que podem ser formalizados em lógica.

5.1 Uso de operadores lógicos

A OWL usa os operadores \sqcup , \sqcap e \neg por meio das palavras reservadas *owl:unionOf*, *owl:intersectionOf* e *owl:complementOf* respectivamente (BECHHOFFER et al., 2004). Nota-se que o conceito de conjunção, disjunção e negação são retirados da teoria dos conjuntos da matemática. Dessa maneira, é essencial simbolizar os sujeitos da linguagem natural que traduzam essa relação.

A **conjunção** determina a interseção entre dois ou mais sujeitos para fazer parte do conjunto, como mostrado nas definições 1, 2 e 4. Dessa forma, qualquer motor



de inferência será capaz de associar uma definição que se enquadre dentro desses dois sujeitos. De forma similar funciona a **disjunção**. Ela determina a união entre dois ou mais sujeitos como mostra a definição 5.

(5) “Fungo ou bactéria” \equiv “Fungo” \sqcup “Bactéria” (Gene Ontology Consortium, 2015).

Dentro desse universo do discurso, qualquer sujeito pertencente a “Fungo” \sqcup “Bactéria” também pertencerá a “Fungo ou Bactéria”. A disjunção é usada pelos motores de inferência identificar futuros sujeitos que podem pertencer a um mesmo conjunto.

Por fim, a **negação** indica que um sujeito é o oposto de outra. Esse operador não é recomendado por Arp, Smith e Spear (2015) sob o argumento que toda definição é possível de ser construída com argumentos positivos, salve algumas exceções como “não-fumante”, por exemplo. Além disso, há o problema das inferências lógicas para a negação como associar todas os sujeitos que não pertencem a fumante ao não-fumante. Ao contrário da interseção e união, a negação é usada para definir apenas um sujeito e não um conjunto deles.

Primeira orientação: os operadores lógicos devem são usados na formalização em lógica sempre que as definições possuem conceitos de interseção ou união entre os sujeitos. Esses conceitos estão associados na língua portuguesa, geralmente, por conjunções coordenativas aditivas e alternativas para os casos de interseção e união (INFANTE; NETO, 2008). Geralmente esses casos se apresentam na língua por meio das palavras “e” e “ou”. O uso da negação deve ser evitado sempre que possível e as definições em linguagem natural devem procurar representarem apenas argumentos positivos. Nos casos onde isso não for possível, observar pelos advérbios de negação da definição. Em ambas as situações, cada caso deve ser observado atentamente pelo profissional da Ciência da Informação.

5.2 Uso de quantificadores

A OWL permite os quantificadores \forall e \exists por meio das palavras reservadas *owl:allValuesFrom* e *owl:someValuesFrom*. Os quantificadores são definidos pela noção de **todo-parte** explicitada por Arp, Smith e Spear (2015) onde (todo sujeito universal possui (algum) sujeito particular.



A teoria da ontologia estabelece os conceitos de definições necessárias e suficientes (AGANETTE; ALMEIDA, 2015). Uma definição em OWL pode ser escrita de forma necessária, ou necessária e suficiente.

Segundo Arp, Smith e Spear (2015) a relação de **necessidade** entre duas definições pode ser mostrada na Definição 6 que estabelece a relação de necessidade entre “dotar de nervo íris” com “nervo da íris”. O fato de dotar de nervo a íris não é condição necessária e suficiente para ser um “nervo da íris”. Ainda é preciso ser um “nervo”.

(6) “*nervo da íris*” \equiv “*nervo*” \sqcap (“*dota de nervo*” \exists “*íris*”) (Gene Ontology Consortium, 2015)

A relação de **necessidade e suficiência** é mostrada na Definição 5 onde ambos os predicados são equivalentes entre si. Ou seja, \forall (Fungo \sqcup Bactéria) \equiv Fungo ou Bactéria.

As relações de **necessidade e suficiência** estão ligadas diretamente aos predicados que relacionam os sujeitos. Elas são usadas pelos motores de inferência para fazer deduções de mundo aberto, ou seja, de possíveis instâncias que podem ser definidas ou permanecem ocultas no modelo de dados.

Na OWL, segundo Horridge (2009), os quantificadores universais devem ser usados com cautela, pois são erros comuns durante a representação do conhecimento. Geralmente o conceito universal já está implícito numa definição. Os quantificadores existenciais são usados por padrão para definir sujeitos e atributos em ontologias.

Segunda orientação: o uso de quantificadores para OWL deve limitar-se às relações de necessidade. Definições e relações entre propriedades que são apenas necessárias usa-se o quantificador existencial \exists enquanto as definições e relações necessárias e suficientes usa-se o quantificador \forall . Por exemplo, ser um “*nervo da íris*” é condição necessária e suficiente para que tenha a definição “*dota de nervo íris*”, mas nem toda definição “*dota de nervo íris*” pode ser considerada um “*nervo da íris*”, já que isso pode ser causado por outros fatores. Logo alguns “*dota de nervo*



íris” são contidos em *“nervos da íris”* e todos os *“nervos da íris”* contém *“dota de nervo íris”*.

5.3 Uso de restrições de cardinalidade

As restrições de cardinalidade são oriundas das linguagens de programação que derivam de conceitos da matemática. Existem três tipos de restrições na OWL: *owl:maxCardinality*, *owl:minCardinality* e *owl:cardinality* que matematicamente são representados pelos símbolos \leq , \geq e $=$ respectivamente. Eles são usados para restringir características relacionadas a sujeitos e predicados por meio de atributos e são usados pelos motores de inferência para classificarem corretamente os sujeitos. A restrição é também semântica e evita ambiguidade no modelo de representação do conhecimento.

A Definição 7 utiliza uma restrição de cardinalidade para instanciar sujeitos singulares. Através das inferências realizadas por meio de restrições de cardinalidade, é possível derivar que uma mulher é múltipara pela quantidade de partos que ela já fez.

(7) *“Múltipara”* \sqsubseteq *“mulher”* \sqcap (*“participa de”* *“parto”* ≥ 3) (FARINELLI et al., 2016).

As restrições de cardinalidade são apenas para atributos. Elas limitam as relações que sujeitos e predicados têm entre si por meio desses atributos. A definição 7, por exemplo, possui o sujeito *“parto”* restringida pela relação *“participa de”* com valor mínimo de 3.

Terceira orientação: a formalização em lógica de restrições de cardinalidade devem ser usadas para restringir os atributos de relacionamento entre sujeitos e predicados. Suas restrições são para qualquer tipo de atributo, sejam eles numéricos, textuais, binários ou temporais.

5.4 Uso de propriedades

Em OWL as propriedades são usadas para ligar dois sujeitos ou relacionar um sujeito com um atributo. Existem inúmeras propriedades na OWL, mas para fins de inferência lógica é importante destacar as seguintes: domínio e imagem; equivalência; simetria; funcional; disjunção mútua; transitividade, inversão,



reflexão e irreflexão (BECHHOFFER et al., 2004). As propriedades estão relacionadas principalmente com os predicados da definição.

5.4.1 Disjunção mútua

A propriedade de disjunção mútua, representada em OWL pela palavra reservada *owl:disjointWith*, é uma restrição para separar sujeitos que são totalmente distintos. Aqui vale ressaltar a diferença entre a disjunção citada na Tabela 1 que tem o conceito de união entre dois conjuntos. Logicamente falando a disjunção mútua pode ser representada com o operador lógico “ou exclusivo” que só é verdadeiro quando dois valores são diferentes.

Os motores de inferência usam a disjunção mútua para detectar inconsistências na ontologia. Tendo dois sujeitos A e B que são mutuamente disjuntos, um sujeito A não pode ser um sujeito B, nem vice-versa.

Quarta orientação: use a propriedade de disjunção mútua para assegurar que sujeitos e predicados não sejam definidos como equivalentes e separá-los logicamente. Do contrário, asserções de mundo aberto podem ser inferidas posteriormente e causar um erro indetectável no modelo de representação do conhecimento. Por exemplo, um modelo com o sujeito “*órgão*” deve ter os sujeitos “*pulmão*”, “*cérebro*” e “*rim*” como subclasses mutuamente disjuntas, pois um “*cérebro*” não pode ser um “*rim*”, nem vice-versa.

5.4.2 Domínio e imagem

As relações de domínio e imagem, representadas em OWL pelas palavras reservadas *rdfs:domain* e *rdfs:range* respectivamente, são derivadas da matemática para representação de funções em planos cartesianos. Em OWL o domínio e imagem funcionam para ligar determinadas propriedades aos sujeitos. É como se a relação entre dois sujeitos fosse uma função de duas variáveis. O domínio representa o ponto *x* da função e a imagem o ponto *y*.

Na prática, o motor de inferência trata domínio e imagem como axiomas e faz inferências baseadas nos sujeitos particulares que possuem as propriedades necessárias para pertencer aos sujeitos universais. Na definição 7 o domínio da



propriedade “*participa de*” seria “*parto*” e a imagem seria uma interseção entre o sujeito “*parto*” e a restrição de cardinalidade ≥ 3 . É importante ressaltar que as propriedades de domínio e imagem não funcionam como restrições. Ou seja, de acordo com essa afirmação, qualquer sujeito que possuir a propriedade “*participa de*” será considerado um “*parto*” o que pode representar um erro de modelagem da ontologia e, portanto, deve ser usado com cautela.

Quinta orientação: as propriedades de domínio e imagem devem ser usadas somente para dizer ao motor de inferência como que os sujeitos se relacionam com as propriedades. As propriedades a serem usadas, no entanto, devem ser exclusivas dos sujeitos representados para que não ocorram erros de modelagem, como o exemplo mostrado acima.

5.4.3 Equivalência

A propriedade de equivalência é representado em OWL pela palavra reservada *owl:equivalentProperty*. A equivalência pode ser usada tanto em nível de atributos quanto em nível de sujeitos. Para o último utiliza-se a palavra reservada *owl:equivalentClass*. O conceito é usado quando dois sujeitos ou atributos possuem todas as características em comum. Também pode-se dizer que dois conceitos equivalentes representam a relação de necessidade e suficiência entre eles.

Os motores de inferência usam o conceito de equivalência para inferir sujeitos universais que possuam a mesma definição e atribuir para outros sujeitos que sejam equivalentes.

Sexta orientação: a equivalência, tanto de atributos quanto de sujeitos, deve ser usada como marcação de definições diferentes que possuam o mesmo significado dentro do modelo de representação. Também usa-se a equivalência para definir logicamente o predicado ao qual trata a definição. Por exemplo, a definição 5 é uma definição para “*fungo ou bactéria*” com os sujeitos “*fungo*” e “*bactéria*”. Logo a relação de equivalência para o sujeito “*fungo ou bactéria*” seria a disjunção entre os sujeitos “*fungo*” e “*bactéria*”.



5.4.4 Simetria

Uma propriedade em OWL pode ser simétrica (*owl:SymmetricProperty*) e assimétrica (*owl:AsymmetricProperty*). Diz-se que uma propriedade é simétrica se a relação mútua de A para B é válida de B para A. Por exemplo, a relação “é adjacente de” que liga dois sujeitos é simétrica, pois se um sujeito A está adjacente a B, então B tem de estar adjacente a A. De maneira análoga são as propriedades assimétricas. No entanto, elas restringem que uma relação de A para B ocorra de B para A.

Os motores de inferência usam a propriedade simétrica para inferir sujeitos com relações que estão explícitas apenas unilateralmente e a propriedade assimétrica para garantir consistência do modelo de representação do conhecimento.

Sétima orientação: Use a propriedade simétrica para garantir que dois sujeitos possuam vínculos entre si de forma igual usando uma única relação. A propriedade assimétrica deve ser usada quando houver a possibilidade de ocorrer uma ruptura no modelo de representação de conhecimento. Nesse caso ela é usada para assegurar que o modelo não representará conceitos errados.

5.4.5 Inversa

Toda relação possui um domínio e imagem, ainda que não definidos explicitamente. A propriedade de inversão, definida em OWL por *owl:inverseOf*, estabelece a relação contrária de um domínio e imagem de uma relação. A Definição 8, por exemplo, possui a relação “contém” que tem como domínio “cavidade pericardial” e imagem “fluido pericardial”. Essa relação, por sua vez, pode ser representada inversamente pela relação “é contido em”. Ou seja, o “fluido pericardial está contido em cavidade pericardial”. Nessa segunda representação o domínio passa a ser “fluido pericardial” e a imagem “cavidade pericardial”. A propriedade de inversão é também simétrica e sua inferência é realizada nos mesmos termos que as propriedades simétricas.

(8) Cavidade pericardial \equiv contém \exists (fluido pericardial) (Gene Ontology Consortium, 2015).



Oitava orientação: a propriedade inversa deve ser usada quando o modelo de representação de conhecimento demandar uma relação simétrica entre dois sujeitos que não podem ser representados por uma única relação. Nesse caso, é necessário a criação de uma segunda relação que seja complemento da primeira e com propriedade inversa.

5.4.6 Funcional

A propriedade funcional, representada em OWL pela palavra reservada *owl:FunctionalProperty*, estabelece que uma relação A só pode ter uma imagem B. Em termos de inferência, se um modelo possui as relações A para B e A para C, logo conclui-se que B e C são equivalentes. Uma forma de restringir essas relações é declarar B e C como sendo mutuamente disjuntos.

Há também a propriedade funcional inversa, representada na OWL pela palavra reservada *owl:InverseFunctionalProperty*. Ela estabelece que uma relação de A para B só pode ter um domínio. Em termos de inferência, se um modelo possui as relações A para C e B para C, logo conclui-se que A e B são equivalentes. Conforme explicado no parágrafo anterior, uma forma de restringir essas relações é declarar A e B como sendo mutuamente disjuntos.

Nona orientação: a propriedade funcional deve ser usada quando o modelo de representação do conhecimento exige que as relações se restrinjam a um único domínio ou imagem. O uso dessa propriedade deve ser usada com o intuito de restringir o uso da relação ou identificar sujeitos equivalentes que usam o mesmo domínio ou imagem. No primeiro caso, a propriedade deve ser usada junto com a disjunção mútua, para garantir que violações não ocorram. No segundo caso a disjunção mútua não deve ser usada.

5.4.7 Transitiva

A propriedade transitiva, representada em OWL pela palavra reservada *owl:TransitiveProperty*, está diretamente ligada com o silogismo hipotético. Uma propriedade transitiva afirma que, se a relação de A para B é válida, de B para C é válida então a relação de A para C também é válida. Por exemplo, a relação “é mais



velho que” é transitiva, pois se o sujeito A é mais velho que o B e o sujeito B é mais velho que o C, então o sujeito A tem que ser mais velho que C.

Os motores de inferências usam a relação de transitividade para identificar sujeitos que se relacionam por meio de propriedades transitivas. Em vias gerais a propriedade transitiva pode ser vista como uma relação de classe e subclasse.

Décima orientação: a propriedade transitiva deve ser usada em relações que possuem a noção de sequência ou subclasse com outros sujeitos. Por exemplo, as relações está *“localizado em”* e *“é anterior a”* são transitivas, pois localizações remetem a subclasses e noções de tempo remetem a sequências.

6 FORMALIZAÇÃO EM LÓGICA NA PRÁTICA: UM EXEMPLO FINAL

A aquisição de conhecimento em ontologias já é um campo bastante consolidado na Ciência da Informação, conforme Almeida (2006). As tarefas de reunir e definir o conhecimento capturado ainda reside na percepção e compreensão do profissional da Ciência da Informação. A formalização de linguagem natural em lógica também segue o mesmo pressuposto.

Dadas as orientações fornecidas na seção 5, cabe ao profissional interpretá-las e usá-las de maneira correta. A linguagem natural possui muitas ambiguidades e nuances que dificulta a tarefa de cobrir todas as situações possíveis que podem existir num modelo de representação do conhecimento. Portanto não é objetivo do presente artigo demonstrar todas as possibilidades de formalização da língua, mas sim ilustrar a sua aplicabilidade em ontologias e os benefícios por trás dessa metodologia.

Considere a seguinte definição em linguagem natural retirada da OntoNeo (FARINELLI et al., 2016):

(9) Transfusão sanguínea é um processo de transfusão de sangue ou componentes do sangue de uma pessoa no sistema circulatório de outra pessoa.

O primeiro passo para a formalização em lógica é identificar os sujeitos e predicados da definição. Como sujeitos têm-se: *“transfusão sanguínea”*, *“processo”*,



“*sangue*”, “*componentes do sangue*”, “*pessoa*” e “*sistema circulatório*”. Como predicados têm-se: “*transfusão*”.

Numa primeira análise nota-se que “*transfusão sanguínea*”, o sujeito que está sendo definido, é uma subclasse de “*processo*” e deve ser assim representado na ontologia. Os demais sujeitos também devem estar representados no modelo para que a formalização lógica seja possível. Os sujeitos “*sangue*”, “*componentes do sangue*” e “*sistema circulatório*” são partes de “*pessoa*”, pois a transfusão só ocorre de uma pessoa para outra, segundo a definição.

Numa segunda análise nota-se os aspectos da relação de “*transfusão*”: (1) ela tem como atributo transfundir “*sangue*” ou “*componentes do sangue*”; (2) ela tem como imagem o “*sistema circulatório*” de uma pessoa. O aspecto (1) pode ser definido em lógica da seguinte maneira: “*transfunde (sangue \sqcup componentes do sangue)*”. Note que o conceito de disjunção aqui representa a união dos conceitos. A transfusão pode ser feita tanto usando sangue ou componentes do sangue. O aspecto (2) pode ser formalizado em lógica da seguinte maneira: “*transfunde para circulação sanguínea*”. Em relação às propriedades de “*transfusão*” tem-se que:

- não é equivalente, pois não há outra relação com o mesmo significado;
 - não é simétrica, pois a transfusão ocorre de “*sangue*” ou “*componentes do sangue*” para o “*sistema circulatório*” e o contrário não é válido;
- pode ser inversa com a relação “*recebe transfusão de*”;
 - não é funcional, pois um “*sangue*” ou “*componentes do sangue*” podem ser transferidos para vários “*sistemas circulatórios*” e um “*sistema circulatório*” pode receber transfusão de vários “*sangues*” ou “*componentes de sangue*”;
- não é transitiva, pois transfusão não possui relações de sequência ou subclasse com outras transfusões.

Depois basta adicionar os quantificadores para que as definições tenham valores de verdade independentemente dos sujeitos. Cada relação deve estar acompanhada de seu respectivo quantificador. Transfundir sangue ou componentes de sangue não é condição necessária e suficiente para ser uma transfusão de sangue. O mesmo é válido para a relação de “*em direção a*” com circulação sanguínea. Logo



ambos os quantificadores devem ser existenciais. O resultado é dividido em duas definições:

(10) transfusão_sanguínea \equiv transfunde \exists (sangue \sqcup componentes do sangue) em_direção_a \exists (sistema circulatório)

(11) pessoa \equiv tem_partes \exists (sangue \sqcap componentes do sangue \sqcap sistema circulatório)

Essa formalização em lógica é, então, incorporada a ontologia por meio da linguagem OWL. O resultado da formalização em lógica de definições formais em linguagem natural ajuda a estruturar uma ontologia mais rigidamente de forma que: (1) controles semânticos sejam feitos para não gerar inconsistências e; (2) inferências sejam feitas e gerem novas asserções resultando em aquisição automático de conhecimento.

7 DISCUSSÃO E TRABALHOS FUTUROS

Esse trabalho teve como objetivo apresentar ao público da Ciência da Informação o uso e as vantagens de modelos como ontologias além de contribuir metodologicamente oferecendo orientações para formalização de linguagem natural em lógica. É sabido que o campo da ontologia aplicada oferece vantagens em relação a outros sistemas de modelagem informacional, pois possui os mesmos recursos de uma taxonomia, de um tesouro e, em adição, oferece que os sujeitos definidos possuam relações entre si. Há, no entanto, outras possibilidades que a ontologia oferece como um artefato computacional. O uso de recursos processáveis por computadores como a lógica agrega às técnicas de representação da informação.

A lógica possui um grande potencial dentro do campo da ontologia aplicada. Ela é, no entanto, um recurso pouco explorado por usuários de ontologias. Procurou-se então desvendar o seu funcionamento, além de explicar os conceitos fundamentais para o entendimento da lógica e como ela atua nas ontologias. Sua potencialidade foi demonstrada por meio de exemplificações de como os motores de inferências agem por trás das ontologias.



Como trabalho futuro, espera-se incluir as orientações apresentadas nesse artigo numa metodologia de construção de ontologias. Um bom método de desenvolvimento de ontologias, garante um padrão e consistência durante a construção desses artefatos. Tais metodologias descrevem cada etapa da construção desde a sua concepção até a sua publicação. Como citado na introdução deste artigo, as metodologias de engenharia de ontologias tratam o desenvolvimento da lógica apenas em níveis básicos. É preciso que esta área seja desenvolvida com mais profundidade com o intuito de aprimorar os benefícios que a lógica produz.

Acredita-se que a melhor maneira de propagar e aperfeiçoar o uso da lógica em sistemas de ontologia aplicada é replicando-a em suas metodologias de construção. Dessa maneira, novas ontologias serão criadas usando esses princípios como base e conseqüentemente, essas ontologias estarão usando todo o potencial que o artefato oferece.

8 AGRADECIMENTOS

Esse trabalho foi financiado pela bolsa CAPES Demanda Social que permitiu ao autor seu desenvolvimento. O conteúdo, no entanto, é de total responsabilidade do autor e não representa, necessariamente, a opinião da CAPES.

REFERÊNCIAS

AGANETTE, E. C.; ALMEIDA, M. B. Avanços teórico-empíricos na comparação entre teoria da ontologia e teoria da terminologia. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 20, n. 44, p. 3–24, 2015.

ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. *Perspectivas em Ciência da Informação*, [S.l.], v. 19, n. 3, p. 242-258, set. 2014. ISSN 19815344.

ARISTÓTELES. *Organon*. 3^a. ed. [S.l.]: Edipro, 2016.

ARP, R.; SMITH, B.; SPEAR, A. D. *Building ontologies with basic formal ontology*. [S.l.]:Mit Press, 2015.

BAADER, F. et al. *Terminological knowledge representation: A proposal for a terminological logic*. 1990.



- BAADER, F.; HORROCKS, I.; SATTLER, U. Description logics. In: Handbook on ontologies. [S.l.]: Springer, 2009. p. 21–43.
- BAADER, F. et al. The Description Logic Handbook: Theory, Implementation and Applications. 2nd. ed. New York, NY, USA: Cambridge University Press, 2010.
- BALLARD, K. E.; COPI, I. M. Study Guide for Copi Introduction to Logic. [S.l.]: Monograph Collection (Matt - Pseudo), 1978.
- BECHHOFER, S. et al. Owl web ontology language reference. W3C recommendation v. 10, n. 02, 2004.
- DIMNET, E. A arte de pensar. [S.l.]: Kírion, 2020.
- FARINELLI, F. et al. Ontoneo: The obstetric and neonatal ontology. 2016.
<https://ontoneo.com/>
- FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. American Association for Artificial Intelligence, 1997.
- FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A.; CORCHO, O. Methodologies and methods for building ontologies. Gómez-PérezA. Fernández-LópezM. CorchoO.(Eds.), Ontological engineering, p. 107–153, 2004
- GENE ONTOLOGY CONSORTIUM. Gene ontology consortium: going forward. Nucleic acids research, Oxford Univ Press, v. 43, n. D1, p. D1049–D1056, 2015.
- GENESERETH, M.; KAO, E. Introduction to logic. Synthesis Lectures on Computer Science, v. 4, n. 1, p. 1-165, 2013.
- GENSLER, H. J. Introduction to logic. [S.l.]: Routledge, 2010.
- GIACOMO, G. D.; LENZERINI, M. Tbox and abox reasoning in expressive description logics. KR, v. 96, n. 316-327, p. 10, 1996.
- GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies, Elsevier, v. 43, n. 5, p. 907–928, 1995.
- GRUBER, T. Ontology. encyclopedia of database systems, ling liu and m. tamer özsü. Springer-Verlag. Cloud Computing Economics: Democratization and Monetization of Services. Journal of Business & Economics Research, v. 7, n. 6, p. 1–11, 2009
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. International journal of human-computer studies, Elsevier, v. 43, n. 5, p. 625–640, 1995.



- HARMELEN, F. V.; LIFSCHITZ, V.; PORTER, B. Handbook of knowledge representation. [S.l.]: Elsevier, 2008.
- HORRIDGE, M.A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.2. [S.l.], 2009.
- INFANTE, U.; NETO, P. C. Gramática da língua portuguesa. [S.l.]: Scipione, 2008. 584 p.
- MCGUINNESS, D. L.; HARMELEN, F. V. et al. Owl web ontology language overview. W3C recommendation, 2004.
- MEHRI, R.; HAARSLEV, V.; CHINAEI, H. Optimizing heuristics for tableau-based owlreasoners.arXiv preprint arXiv:1810.06617, 2018.
- MENDONÇA, F. M. Ontoforinfoscience: metodologia para construção de ontologias pelos cientistas da informação-Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA). Tese (Doutorado), 2015.
- MENDONÇA, F. M.; ALMEIDA, M. B. Ontoforinfoscience: A detailed methodology for construction of ontologies and its application in the blood domain. Brazilian Journal of Information Science, Universidade Estadual Paulista-UNESP, Programa de Pós-Graduação em Ciência da Informação, v. 10, n. 1, 2016.
- MUNN, K.; SMITH, B. Applied ontology: an introduction. [S.l.]: Walter de Gruyter, 2013. v. 9.
- NOY, N. F.; MCGUINNESS, D. L. et al. Ontology development 101: A guide to creating your first ontology. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA, 2001.
- PATEL-SCHNEIDER, P. F. et al. Owl web ontology language semantics and abstract syntax. W3C recommendation, v. 10, 2004.
- SANTOS, M. T. d. Estudo do processo de apropriação da ontologia pela Ciência da Informação no Brasil. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2014.
- SEPPÄLÄ, S.; RUTTENBERG, R. Survey on defining practices in ontologies: Report in the international workshop on definitions in ontologies. ICBO 2013, 2013.
- SEPPÄLÄ, S.; SCHREIBER, Y.; RUTTENBERG, A. Textual and logical definitions in ontologies, Lecture Notes, p. 35, 2014.
- SEPPÄLÄ, S.; RUTTENBERG, A.; SMITH, B. Guidelines for writing definitions in ontologies. Ciência da Informação, v. 46, n. 1, 2017.
- SHEARER, R.; MOTIK, B.; HORROCKS, I. Hermit: A highly-efficient owl reasoner. In: OWLED. [S.l.: s.n.], 2008. v. 432, p. 91.



- SIPSER, M. Introduction to the theory of computation. ACM Sigact News, ACM New York, NY, USA, v. 27, n. 1, p. 27–29, 1996.
- SIRIN, E. et al. Pellet: A practical owl-dl reasoner. Web Semantics: science, services and agents on the World Wide Web, Elsevier, v. 5, n. 2, p. 51–53, 2007.
- SMITH, B. Ontology and Information Science. 2003.
<https://philpapers.org/rec/SMIOAI>
- SMITH, B. et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology, Nature Publishing Group, v. 25, n. 11, p. 1251, 2007
- SMITH, B.; CEUSTERS, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied ontology, IOS Press, v. 5, n. 3-4, p. 139–188, 2010.
- SMITH, B. et al. Basic Formal Ontology 2.0 Draft Specification and User's Guide. 2014.
- SUÁREZ-FIGUEROA, M. C. et al. Ontology engineering in a networked world. [S.l.]: Springer Science & Business Media, 2012.
- SUPPES, P. Introduction to logic. [S.l.]: Courier Corporation, 2012.
- TSARKOV, D.; HORROCKS, I. Fact++ description logic reasoner: System description. In: SPRINGER. International Joint Conference on Automated Reasoning. [S.l.], 2006. p. 292–297.
- USÓ-DOMÉNECH, J. L.; NESCOLARDE-SELVA, J. A.; GASH, H. Universe of discourse and existence. Mathematics, Multidisciplinary Digital Publishing Institute, v. 6, n. 11, p. 272, 2018
- VICKERY, B. C. Ontologies. Journal of information science, Sage Publications, v. 23, n. 4, p. 277–286, 1997.

